RESEARCH ARTICLE

Automated Artifact Rejection Algorithms Harm P3 Speller Brain-Computer Interface Performance

David E. Thompson[a], Md. Rakibul Mowla[a], Katie J Dhuyvetter[a], Joseph W. Tillman[a], and Jane E. Huggins[b]

[a]Brain and Body Sensing (BBS) Lab, Mike Wiegers Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS, USA;
[b]Direct Brain Interface Lab, Department of Physical Medicine and Rehabilitation and Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

ARTICLE HISTORY
Compiled October 5, 2021

ABSTRACT
Brain-Computer Interfaces (BCIs) have been used to restore communication and control to people with severe paralysis. However, non-invasive BCIs based on electroencephalogram (EEG) are particularly vulnerable to noise artifacts. These artifacts, including electro-oculogram (EOG), can be orders of magnitude larger than the signal to be detected. Many automated methods have been proposed to remove EOG and other artifacts from EEG recordings, most based on blind source separation. This work presents a performance comparison of ten different automated artifact removal methods. Unfortunately, all tested methods substantially and significantly reduced P3 Speller BCI performance, and all methods were more likely to reduce performance than increase it. The least harmful methods were titled SOBI, JADER, and EFICA, but even these methods caused an average of approximately ten percentage points drop in BCI accuracy. Possible mechanistic causes for this empirical performance deduction are proposed.

KEYWORDS
Brain-computer interfaces, P300 speller, artifacts rejection, physiological signals, signal processing.

1. Introduction

Brain-Computer Interfaces (BCIs) are an assistive technology tool designed to allow people with the most significant movement impairments to communicate and control their environment [1]. For these individuals, particularly those with locked-in syndrome who are without other means of voluntary control, BCIs represent the only hope of autonomy.

While there are many competing technologies and techniques, the most commonly-used BCI to date is the P300 or P3 Speller based on electro-encephalogram (EEG), first proposed by Farwell and Donchin [2]. The P3 Speller BCI has been shown to be suitable for in-home use by people with paralysis, including a research scientist who used the BCI to maintain employment [3, 4] and another person who used it to control wheelchair tilt [5].

However, substantial concerns have been raised regarding the limits of the usefulness of the P3 Speller. Visual P3 Spellers tend to perform poorly if the user cannot fixate on the letter of interest [6]. Other groups have demonstrated that redesigning the interface allows communication without fixation [7], though with a substantial performance penalty. Our prior work has also demonstrated the importance of latency variation for some users [8], and offline

David Thompson. Email: davet@ksu.edu
Md Rakibul Mowla. Email: rakibulmowla@ksu.edu/rakib.raju05@gmail.com
Jane Huggins. Email: janeh@med.umich.edu

analysis has independently confirmed the efficacy of correction for latency differences between tasks [9].

This paper focuses on another potential confound with these visual interfaces: eye movements and blinking. Eye movements, including blinking, produce artifacts in EEG recordings that are orders of magnitude larger than the P3 potential. There are several mechanisms by which blinking might adversely affect P3 Speller operation: 1) Blinks could cause stimuli to be missed or perceived at the wrong time, 2) Electro-oculogram (EOG) artifacts including those generated by blinking could obscure the P3 signal, and 3) Instruction to withold blinking can cause reductions in the P3 amplitude [10], making detection more difficult. The first problem is perhaps the easiest to address, as detecting blinks from EEG is relatively simple; therefore this paper will focus on the other two mechanisms. We argue that the second and third problems could be addressed by the same solution: namely, the removal of blink artifacts from the EEG signal. If an automated system could remove these EOG artifacts, then experimenters could refrain from giving blink instructions that might interfere with the amplitude of the P3 signal.

Many groups have developed methods for removing EOG artifacts, including those from blinking, from online and offline EEG recordings. Many of these methods are based on blind source separation (BSS) techniques. Such methods attempt to reconstruct the underlying EEG signal by subtracting out the estimated component(s) related to EOG activity. Examples include [11, 12, 13, 14, 15, 16, 17, 18, 19, 20] among many others.

Rather than propose yet another new automated artifact removal (AAR) algorithm, our group set out to answer two research questions. The first is whether AAR algorithms could improve the performance of the P3 Speller. The second question is which existing method would produce the best P3 Speller performance.

To our knowledge, most of the AAR methods we selected have not been used previously in the setting of the P3 Speller. The exception is [18], which reported a slight reduction in accuracy (69% to 68%) in an auditory P3 Speller task. We thus believe this to be the first comparison of AAR methods for the P3 Speller, with the exception of our prior work reported as abstracts at BCI conferences [21, 22].

2. Methods

2.1. *Data Source*

As we wished to compare many methods, we chose to compare methods with offline analysis rather than create multiple online experiments. As our data source, we selected a large dataset from our previous work [8, 23]. The experimental protocol consisted of 9 sentences typed with correction over three separate visits. The protocol was begun by 40 participants, including 25 men (age: 45±21 years, mean±std. dev) and 15 women (age: 42±20 years). Of these participants, eight men and four women were diagnosed with Amyotrophic Lateral Sclerosis, while five men and one woman were diagnosed with Muscular Dystrophy. Four participants did not return for a second session, while one dropped out during the second session. We also dropped seven sentences that were not completed, and three for which one of the AAR methods did not work. The final dataset includes 320 testing files and 40 training files. For additional demographic information, please see our previous publications. All data were gathered under approved protocols under the auspices of the University of Michigan's Institutional Review Board (IRB). The current analysis is covered by Kansas State University's IRB, under exempt protocol #7516.

2.2. *Original Experimental Setup*

As stated, participants operated the BCI over three separate visits. During the first visit, a training file was generated using the 19-character phrase "A QUICK BROWN FOX" (spaces included). During each visit (including the first), participants copied three sentences into three separate environments. Participants were instructed to correct errors through a backspace option presented in the P3 Speller BCI. Consequently, the number of characters per sentence varies by participant, ranging from 23 (the sentence length) to 77.

In a copy spelling with correction task, determining the target character is not straightforward. Our baseline assumption was that the target character was the next character in the sentence, unless errors were present in the text. If errors were present, the backspace was the target character. However, we supplemented this assumption with two sources of information. First, after each sentence we asked each participant if he or she focused on incorrect characters, and we recorded any reported deviations. Second, we inspected each sentence individually to see if deviations from the expected pattern were present. An example case: one participant missed a single letter, then typed the following six characters correctly, before verbally acknowledging the missed character and backspacing back to the point of the original error. In such cases, the target character was inferred from available information (in the example, the "correct" characters until the point where backspacing began, then backspaces from that point). The determination of target characters was completed during the analysis in [8], and was not revisited for this work.

2.3. *Recordings and P3 Detection*

The original recordings were performed with Electro-Cap caps with wet electrodes at 10-20 locations F3, Fz, F4, T7, C3, Cz, C4, T8, CP3, CP4, P3, Pz, P4, PO7, Oz, and PO8. Reference and ground were the right and left mastoids, respectively. The potentials were filtered and digitized by a 16-channel g.USBamp system using hardware passband and 60Hz notch filters.

For detection of the P3 signal, we chose the popular and simple least squares classification method. While other methods including step-wise linear discriminant analysis [24, 25] and more complex classifiers such as those based on support-vector machines [26] have produced good results in the literature, we wished to focus the paper on the effect of AAR methods rather than the classifier. In our opinion and experience, least-squares classification provides a baseline estimate of the performance of most linear classifiers.

2.4. *AAR Algorithms*

We tested nine EOG-removal algorithms in the publicly-available EEGLAB toolbox for MATLAB (Mathworks), version 13.6.5b. Eight methods were from the AAR plugin [11], under the "Correct EOG using BSS" dropdown option. These methods are titled SOBI [12], IWASOBI [13], EFICA [14], MULTICOMBI [15], FCOMBI [16], JADER [17], BSSCCA [19], and PCA. We also used the MARA [18] EEGLAB plugin, configured to fully-automated mode.

Finally, we included results from Mr. Mowla's previously published AAR method [20] for comparison, bringing us to a total of ten AAR methods.

Details on each method are available in the corresponding cited publications. Importantly, all methods were left at default settings. The consequences of this decision will be discussed in the Limitations section, but we believed this to be a likely use case for a new researcher incorporating a "fully automated" method.

2.5. *Framework for Comparing AAR Algorithms*

Each AAR method was applied to each file, including the training files. The training files were used to build per-participant classifiers, one for each AAR method. These classifiers were applied to each participant's cleaned test data. In all cases, a classifier trained and tested on data cleaned with the same method (e.g. training on a training file cleaned by SOBI, testing on a testing file cleaned by SOBI). The performance of each method was compared to using raw train and test data.

2.6. *Metrics and Statistical Testing*

We have previously argued [27] that the most useful metric for the P3 speller is BCI Utility [28] because it enables comparisons between studies. In this case, all comparisons are between methods applied to the same datafile. Therefore, unlike comparisons between different studies, each method has the same amount of data and time available. Since the timing is consistent between the methods, BCI Utility is a direct translation of accuracy. We have reported both metrics for consistency with our prior recommendations; as the information within is equivalent, the reader may focus on whichever metric is more familiar. All metrics are calculated on a per-file (per-sentence) basis. The average change for each metric is presented in the Results section.

Our test for statistical significance is derived from the question "Is this method more likely than not to help (or hurt) performance, as compared to not using it?" In order to answer this question, we first noted the most common performance change, and then considered how likely that outcome was as compared to the other possible outcomes. For example, if the method decreased performance, we then compared the number of sentences with a performance decrease to the number of sentences that had either better or equivalent performance. The files were then considered independent Bernoulli random variables; the sum of the performance should then follow a binomial distribution. The MATLAB function "binofit" was used to develop confidence bounds for the likelihood of a change in the indicated direction. A lower confidence bound above 50% indicated the method was not changing performance randomly, but instead had a statistically significant bias toward either better or worse performance. Given our number of comparisons, we used Bonferroni correction when calculating the confidence bounds ($\alpha = 0.005$, or 0.05/10 methods).

When testing for significance using this approach, we used the accuracy rather than Utility numbers. The sign of the change is equivalent except for a few cases where the original Utility was zero due to an accuracy below 50%; we still wished to capture changes in performance, positive or negative, for these files.

3. Results

The mean and median changes in accuracy and BCI Utility are shown in Table 1. Given the number of participants, files, and methods, individual results are not tabulated. As can be seen from the table, the performance changes are uniformly negative, and substantial.

Table 1. Changes in accuracy and BCI Utility (mean and median) for each method. Accuracy changes are reported in percentage points, not relative change. BCI Utility was calculated using the formulas for corrected characters per minute.

Method	Mean ΔAcc (%)	Med. ΔAcc (%)	Mean ΔUtility (char/min)	Med. ΔUtility (char/min)
BSSCCA	-17.4	-14.9	-0.19	-0.15
EFICA	-11.5	-7.55	-0.14	-0.08
FCOMBI	-19.6	-15.7	-0.23	-0.16
IWASOBI	-20.5	-16.3	-0.24	-0.17
JADER	-9.2	-6.1	-0.11	-0.07
MARA	-70.3	-77.8	-0.47	-0.48
MULTICOMBI	-22.6	-19.4	-0.27	-0.19
PCA	-22	-18.5	-0.23	-0.17
SOBI	-9.7	-6.7	-0.12	-0.08
MOWLA	-23.3	-22.6	-0.28	-0.24

Figures 1 and 2 display box plots for each method's change in accuracy and Utility, respectively. The plots were made with MATLAB's "boxplot" command. As shown in the figures, only three methods (SOBI, JADER, and EFICA) include zero change in their central quartiles.

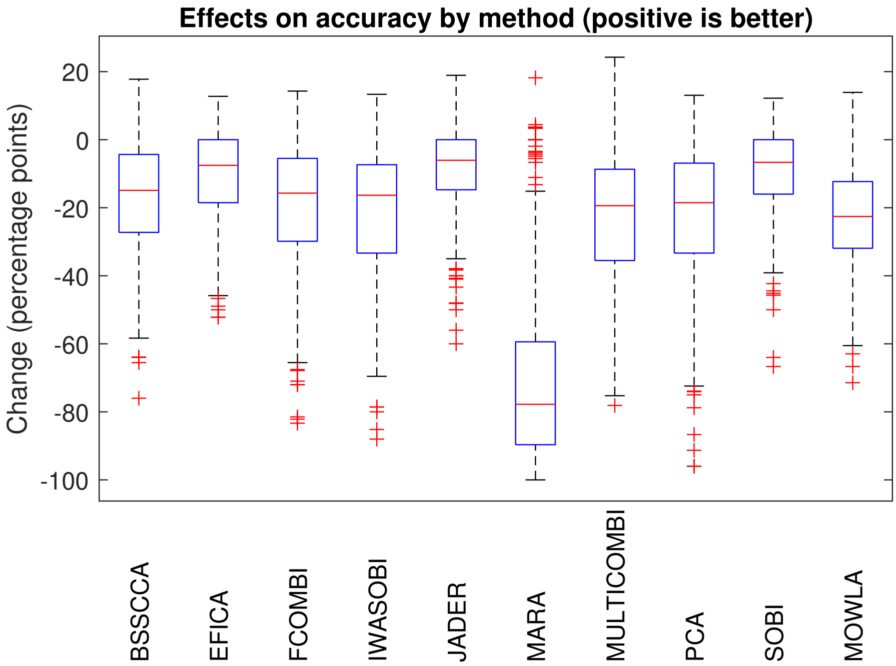

Figure 1. Boxplot of change in accuracy when applying each method (negative change indicates reduced performance). Red line indicates median, blue box denotes 25th and 75th percentiles. Whiskers enclose all data not marked as an outlier by MATLAB's default method; asterisks indicate outliers.

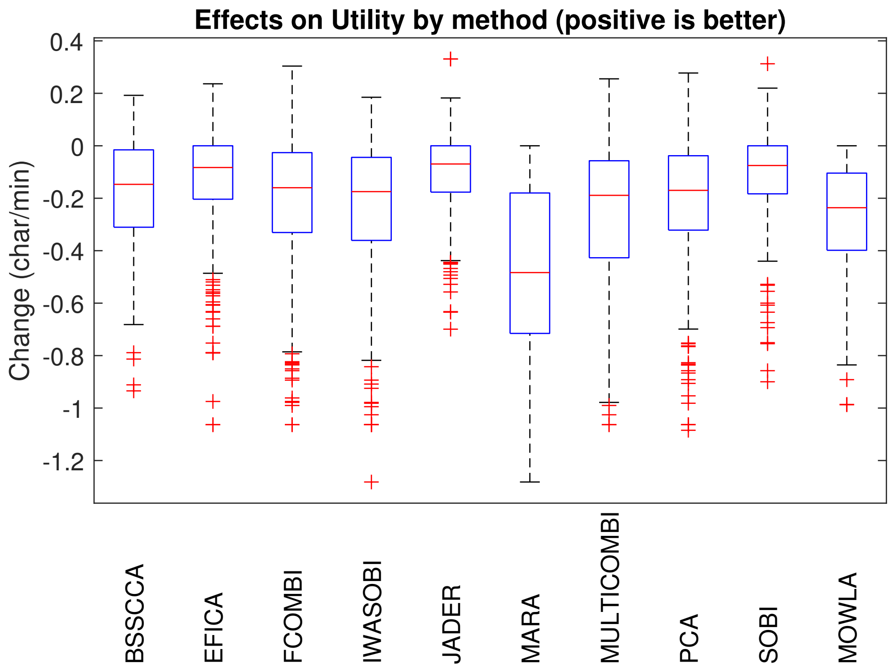

Figure 2. Boxplot of change in BCI Utility when applying each method (negative change indicates reduced performance).

The number of sentences for which performance decreased is listed for each method in Table 2. This table also lists the estimated percent chance that the method decreases performance, along with the corresponding confidence intervals. A confidence interval with lower bound above 0.50 indicates the method is statistically more likely to hurt performance than improve it or leave it unchanged. As shown in the table, all AAR methods were statistically more likely to hurt performance than any other outcome.

Table 2. Number of "harmed" sentences (out of 320), along with estimates of probability and corresponding confidence interval (C.I.).

Method	Harmed Sentences	Estimated Probability	C.I. Lower Bound	C.I. Upper Bound
BSSCCA	267	0.83	0.77	0.89
EFICA	232	0.73	0.65	0.79
FCOMBI	269	0.84	0.78	0.89
IWASOBI	277	0.87	0.80	0.91
JADER	231	0.72	0.65	0.79
MARA	305	0.95	0.91	0.98
MULTICOMBI	281	0.88	0.82	0.92
PCA	274	0.86	0.79	0.91
SOBI	229	0.72	0.64	0.78
MOWLA	304	0.95	0.91	0.98

Finally, the individual, sentence-level results are presented in Figures 3 and 4. These figures show the performance for each sentence, for each AAR method, plotted against the raw (no AAR) performance. Figure 3 uses accuracy, Figure 4 shows BCI Utility. The overall negative effect on performance is clearly visible in the figures. These figures further demonstrate the lack of pattern in the results. In our past experience, new techniques have at times shown promise for participants with poor online performance but no gain for those with good online performance. Unfortunately, in this case it can be seen that no AAR method shows consistent improvement at any initial level of performance.

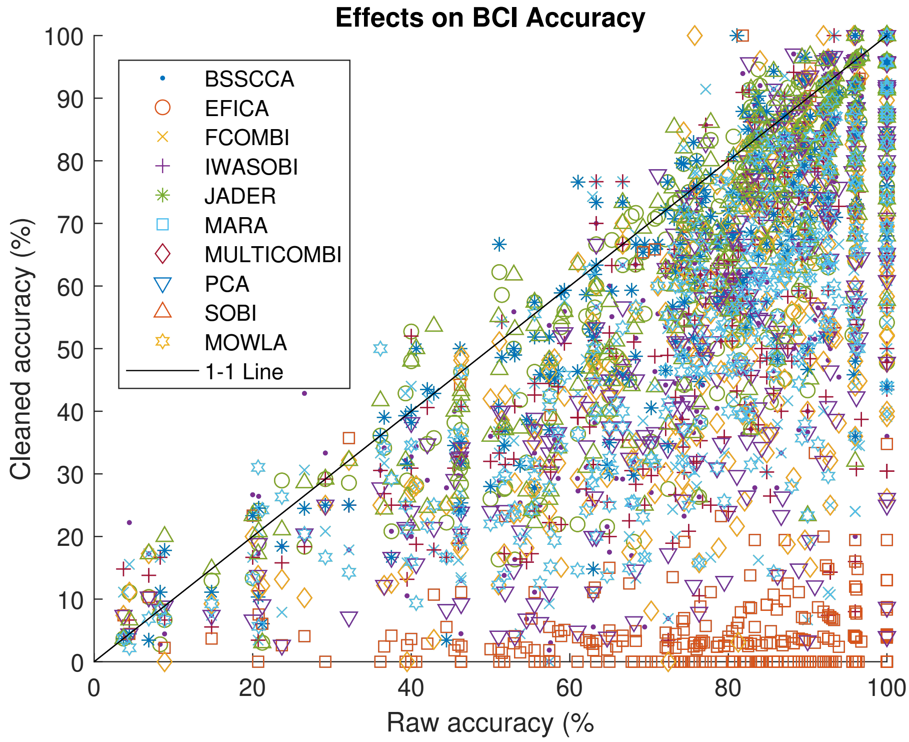

Figure 3. Cleaned vs Raw Accuracy for ten AAR methods, in percent. The 1-1 Line represents no change in performance. Any icons above the line are performance increases, while those below the line represent performance decreases.

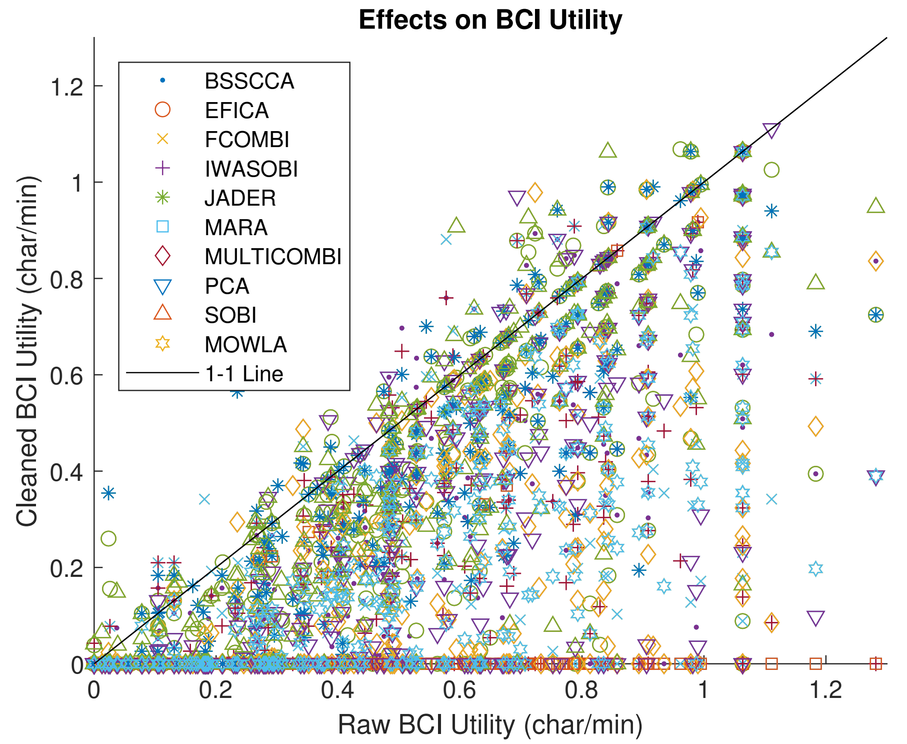

Figure 4. Cleaned vs Raw BCI Utility for ten AAR methods, in corrected characters per minute. The 1-1 Line represents no change in performance. Any icons above the line are performance increases, while those below the line represent performance decreases.

4. Discussion

The primary finding from the study is straightforward: for the vast majority of participants, the use of BSS-based AAR algorithms led to substantially reduced BCI performance. However, the consequences of this finding depend on the reason or mechanism behind this performance penalty. Here, we will propose several possible mechanisms for this penalty; formal investigation of each of these mechanisms is out of the scope of the current paper and will be left for future work.

The first possible mechanism is a source renumbering or lack of source consistency. Among BSS techniques, there is little guarantee that the same source present in different data will be numbered similarly. This mechanism could be ruinous to the performance of classification based on the individual components. However, since all the methods we used re-combined the components back into a time-domain signal prior to classification, we do not believe this issue to be the cause of our findings.

The second possible mechanism is P3 rejection: where the AAR methods incorrectly label some or all of the P3 signal as noise. Like blink and EOG artifacts, the P3 is a spatially broad, low-frequency potential. While the spatial distribution of P3 is different to that of EOG, it seems possible that BSS methods designed to find EOG may incorrectly identify the P3 as an artifact. Some corroboration of this mechanism is provided by a reduced P3 amplitude that we have observed after applying the AAR methods. However, the standard deviation of the EEG signals is also reduced by an equal or greater measure, implying that the detectability or signal-to-noise ratio should not be greatly affected by this mechanism. If our initial assessment is incorrect, and this is a substantive issue, the development of P3-specific AAR algorithms would be indicated.

A third possible mechanism is based on properties of linear algebra relevant to our classifier. All of the methods investigated are BSS-based methods. BSS methods are fundamentally similar in that they identify sources, label some as noise, and reconstruct the signal after discarding some number of sources. The resulting cleaned signal matrix is by definition lower rank than the original signal. If the original signal was rank N originally, and M components were discarded, the resulting signal matrix is rank N-M. Since the signal matrix is not full-rank, the numerical inversion performed in linear regression becomes less stable and more sensitive to small noise perturbations. Our initial investigations lend some credence to this possible mechanism; the condition number of the matrix inverted during linear regression increases by a factor of approximately 10^5 after cleaning. If this is the mechanism causing the reduction in classification performance, then reducing the number of reconstructed channels may be a solution to the observed issue. However, it should be noted that our initial investigations into step-wise linear discriminant analysis, which should be less vulnerable to this particular mechanism, did not show any notable difference in performance penalty.

A fourth possible mechanism is that the EOG was being inadvertently used for classification. While the participants should have been moving their eyes only between characters rather than during the presentation of stimuli, participants did fixate on the characters and EOG is large relative to the P3. The temporary EOG from fixation, which should have resolved before flashing began, would not be synchronized with target presentations and should, therefore, not have affected classification on average. Stereotyped saccades immediately after target stimuli, however, could theoretically produce EOG consistent enough to affect classification. We did not observe such movements during recordings, and these movements seem unlikely to have been performed by large numbers of participants. Nevertheless, further investigation will be necessary to rule out EOG-based classification on at least a subset of participants.

Our fifth and final possible mechanism is perhaps the most concerning. Could the participants, knowingly or inadvertently, have been timing their blinks such that they were using blinks to control the "brain-computer" interface? Certainly, in our observations, participants are likely to withhold blinking while waiting for an upcoming stimulus. Like eye movements, blinks produce EOG that is orders of magnitude larger than the P3. A classifier could easily latch onto such a large signal, provided the timing is tight enough, and the average person's blink reaction time is easily less than the classification window.

If either the fourth or fifth mechanism is causing the performance penalty, then the AAR methods are really only removing "cheating" that a fully paralyzed person could not perform. Thus, we argue that, if some combination of these mechanisms is the cause of the penalty, all labs should include EOG rejection even though it will hurt the apparent performance of their BCI. The labs would gain in return, however, the knowledge that their performance estimates are far closer to that achievable by those without voluntary eye or blink control. Given the throughput of current BCI systems as compared to eye trackers and other assistive technology [7], including single-switch scanning devices, people with voluntary eye and blink control already have better options than a BCI.

We did find some evidence that at few AAR methods may be helpful for a small subset of participants. These individuals may play a key role in elucidating the underlying mechanisms at play; alternatively, they may be exemplars of an entirely different mechanism.

We believe the MARA publication [18] to be the only prior P3 speller investigation with an AAR method. They, too, reported a reduction in performance but it was much smaller than that reported here (roughly one percentage point, from 69 to 68). Several differences in approach might explain the larger reduction in our study. First, they used auditory P3 speller data, which likely had less EOG artifact. Second, they used MARA's supervised rejection mode rather than the fully automated mode used in our study. Third, their EEG setup had significantly more channels – 64 to our 16 – which would be expected to increase the usefulness of BSS methods. Fourth, their pre-processing and classification results were considerably different.

4.1. *Limitations*

This work is a retrospective analysis of existing data which was collected for a separate purpose. While we believe the comparison between methods is fair, the work could be strengthened through replication by other laboratories or on a new dataset. We acknowledge that online use may also lead to different outcomes than offline analyses, though the present results are not encouraging.

The usefulness of BSS methods may also depend on the number of channels. This study used a relatively small number of channels (16), whereas our ongoing data collection uses 64 channels. In future work, we hope to investigate if any of these methods are helpful in a higher channel-count setting.

Currently, we are investigating possible mechanisms by which the observed lowering of performance can be explained. The results of these investigations are expected to direct future research effort toward minimizing the observed performance penalties.

Finally, it is possible that one or more methods could benefit substantially from parameter tuning. While the AAR toolbox uses a published approach [11] to automate the process, there are still parameters that could possibly improve performance if optimized. MARA's documentation, in particular, implies that fully-automated operation is expected to produce sub-optimal performance. Our primary result should be interpreted in light of this limitation.

5. Conclusion

While the SOBI, JADER, and EFICA methods were the least harmful, all BSS-based algorithms used in this study substantially decreased P3 Speller performance. This is an empirical finding, but the effect is not small. We have proposed candidate mechanisms for this performance decrease which are at least *a priori* reasonable. However, the proper response of the BCI community to this empirical performance reduction depends heavily on the mechanism that causes it. Further investigation into the mechanisms behind this performance penalty are warranted and underway. In the meantime, researchers are cautioned that the choice of including EOG rejection algorithms can cause larger-than-expected differences in overall system throughput.

Acknowledgement

The data presented here were collected under National Institute of Child Health and Human Development (NICHD), the National Institutes of Health (NIH) under Grant R21HD054697 and by the National Institute on Disability and Rehabilitation Research (NIDRR) in the Department of Education under Grant H133G090005 and Award Number H133P090008. Data collection and early analysis were supported by the National Science Foundation (NSF) Graduate Student Research Fellowship under Grant DGE0718128. The present analysis was supported by Kansas State University (KSU) faculty startup funds, undergraduate research support from the K-State Department of Electrical and Computer Engineering, and the NSF under Award 1910526. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the respective funding agencies.

References

[1] Wolpaw JR, Birbaumer N, McFarland DJ, et al. Brain-computer interfaces for communication and control. Clinical Neurophysiology. 2002 Jun;113(6):767–791.

[2] Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology. 1988 Dec;70(6):510–523. Available from: http://www.ncbi.nlm.nih.gov.proxy.lib.umich.edu/pubmed/2461285.

[3] Holz EM, Botrel L, Kaufmann T, et al. Long-term independent brain-computer interface home use improves quality of life of a patient in the locked-in state: a case study. Archives of Physical Medicine and Rehabilitation. 2015 Mar;96(3 Suppl):S16–26.

[4] Sellers EW, Vaughan TM, Wolpaw JR. A brain-computer interface for long-term independent home use. Amyotrophic Lateral Sclerosis. 2010 Oct;11(5):449–455. WOS:000283069200007.

[5] Thompson DE, McCann MT, Huggins JE. Controlling wheelchair tilt with a brain-computer interface for in-home use. Presented at the 4th International BCI Meeting, Asilomar, California. 2010.

[6] Brunner P, Joshi S, Briskin S, et al. Does the 'P300' speller depend on eye gaze? Journal of Neural Engineering. 2010 Oct;7(5). WOS:000282097500013.

[7] Treder MS, Blankertz B. (C)overt attention and visual speller design in an ERP-based brain-computer interface. Behavioral and Brain Functions. 2010 May;6(1):28.

[8] Thompson DE, Warschausky S, Huggins JE. Classifier-based latency estimation: a novel way to estimate and predict BCI accuracy. Journal of Neural Engineering. 2013 Feb;10(1):016006. Available from: http://iopscience.iop.org/1741-2552/10/1/016006.

[9] Iturrate I, Chavarriaga R, Montesano L, et al. Latency correction of event-related potentials between different experimental protocols. Journal of Neural Engineering. 2014 Jun;11(3):036005. Available from: http://iopscience.iop.org/1741-2552/11/3/036005.

[10] Verleger R. The instruction to refrain from blinking affects auditory P3 and N1 amplitudes. Electroencephalography and Clinical Neurophysiology. 1991 Mar;78(3):240–251. Available from: http://www.ncbi.nlm.nih.gov.proxy.lib.umich.edu/pubmed/1707797.

[11] Gomez-Herrero G, Clercq WD, Anwar H, et al. Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel. In: Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006; Jun.; 2006. p. 130–133.

[12] Belouchrani A, Abed-Meraim K, Cardoso J, et al. A blind source separation technique using second-order statistics. IEEE Transactions on Signal Processing. 1997 Feb;45(2).

[13] Tichavský P, Doron E, Yeredor A, et al. A computationally affordable implementation of an asymptotically optimal BSS algorithm for ar sources. In: 2006 14th European Signal Processing Conference; Sep.; 2006. p. 1–5.
[14] Koldovsky Z, Tichavsky P, Oja E. Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the CramÉr-Rao Lower Bound. IEEE Transactions on Neural Networks. 2006 Sep;17(5):1265–1277.
[15] Tichavsky P, Koldovsky Z, Yeredor A, et al. A Hybrid Technique for Blind Separation of Non-Gaussian and Time-Correlated Sources Using a Multicomponent Approach. IEEE Transactions on Neural Networks. 2008 Mar;19(3):421–430.
[16] Gomez-Herrero G, Koldovsky Z, Tichavsky P, et al. A Fast Algorithm for Blind Separation of Non-Gaussian and Time-Correlated Signals. In: in Proceedings of the 15th European Signal Processing Conference, Poznan, Poland; 2007. p. 5.
[17] Cardoso JF, Souloumiac A. Blind beamforming for non-Gaussian signals. In: IEE proceedings F (radar and signal processing); Vol. 140. IET; 1993. p. 362–370.
[18] Winkler I, Brandl S, Horn F, et al. Robust artifactual independent component classification for BCI practitioners. Journal of neural engineering. 2014;11(3):035013.
[19] Clercq WD, Vergult A, Vanrumste B, et al. Canonical Correlation Analysis Applied to Remove Muscle Artifacts From the Electroencephalogram. IEEE Transactions on Biomedical Engineering. 2006 Nov;53(12):2583–2587.
[20] Mowla MR, Ng SC, Zilany MSA, et al. Artifacts-matched blind source separation and wavelet transform for multichannel EEG denoising. Biomedical Signal Processing and Control. 2015 Sep;22:111–118. Available from: http://www.sciencedirect.com/science/article/pii/S1746809415001111.
[21] Tillman JW, Huggins JE, Thompson DE. Blink Artifact Rejection Reduces P3 Speller Accuracy but May Prevent Unintended Blink-Based Control. In: Proceedings of the 6th International Brain-Computer Interface Meeting; Asilomar, California; 2016. p. 49.
[22] Dhuyvetter KJ, Huggins JE, Thompson DE. Nine automatic artifact rejection algorithms all decrease P3 speller accuracy. In: Proceedings of the 7th International Brain-Computer Interface Meeting; Asilomar, California; 2018. p. 217–218.
[23] Thompson DE, Baker JJ, Sarnacki WA, et al. Plug-and-play Brain-Computer Interface Keyboard Performance. In: 2009 4th International IEEE/EMBS Conference on Neural Engineering; New York. IEEE; 2009. p. 426–428. WOS:000274650800106.
[24] Krusienski DJ, Sellers EW, McFarland DJ, et al. Toward enhanced P300 speller performance. Journal of Neuroscience Methods. 2008 Jan;167(1):15–21. Available from: http://www.ncbi.nlm.nih.gov.proxy.lib.umich.edu/pubmed/17822777.
[25] Krusienski DJ, Sellers EW, Cabestaing F, et al. A comparison of classification techniques for the P300 Speller. Journal of Neural Engineering. 2006 Dec;3(4):299–305. Available from: http://www.ncbi.nlm.nih.gov.proxy.lib.umich.edu/pubmed/17124334.
[26] Joachims T. Making large-scale svm learning practical. Technical Report; 1998. Report no.:.
[27] Thompson DE, Quitadamo LR, Mainardi L, et al. Performance measurement for brain–computer or brain–machine interfaces: a tutorial. Journal of Neural Engineering. 2014 Jun;11(3):035001.
[28] Dal Seno B, Matteucci M, Mainardi LT. The utility metric: a novel method to assess the overall performance of discrete brain-computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2010 Feb;18(1):20–28. Available from: http://www.ncbi.nlm.nih.gov.proxy.lib.umich.edu/pubmed/20064766.